To Pay or Not to Pay; That is the Question

by
Geraldine Mooney
Mathematica Policy Research Inc.

Lee Giesbrecht
Bureau of the Census

Carolyn Shettle
National Science Foundation

Presented at the 48th Annual Conference of the American Association for Public Opinion Research, St. Charles, Illinois, May 20-23, 1993.

To Pay or Not to Pay; That is the Question

by
Geraldine Mooney
Mathematica Policy Research Inc.

Lee Giesbrecht
Bureau of the Census

Carolyn Shettle
National Science Foundation

Introduction

This paper reports the findings of the 1992 Postcensal pretest conducted by the Bureau of the Census for the National Science Foundation (NSF). Using a four-treatment group experimental design, this pretest examined the impact of questionnaire length and the use of incentives on four outcomes in a mail survey with telephone and in-person follow-ups: 1) cooperation rate; 2) reducing a science and engineering bias; 3) data quality; and 4) "commitment."

Background

The National Science Foundation (NSF) is responsible for collecting information on scientists and engineers in the United States. The Postcensal Survey, conducted by the Bureau of the Census, is one of three major surveys sponsored by NSF to meet this responsibility. This longitudinal study, operates on a 10-year cycle, using the respondents of the Decennial Census long form as its sample frame. During the 1980s low response rates were a serious problem, as exemplified by the 70.6 response rate achieved by the 1982 Postcensal, the baseline survey for a subsequent panel conducted biennially throughout the 1980s.

With better funding for the 1990s, NSF began a redesign of this mail survey, recently renamed the 1993 National Survey of College Graduates. Data collection procedures were expanded, and an 80 percent minimum response rate goal was established. The questionnaire was reformatted to be more "user-friendly" and to minimize a recognized science and engineering (S&E) bias. Part of the redesign, however, included adding new items which more than doubled the level of respondent burden and questionnaire length.

Compared to Postcensal questionnaires of the 1980s, which ranged from six to eight pages, and averaged between 10 to 15 minutes to complete, the 1992 Postcensal pretest questionnaire was a 20-page booklet that took nearly 30 minutes to complete. Although optimal length estimates, with respect to response rates, vary for mail questionnaires of low salience, they tend to range from between two to six pages (Eros, 1970, Sudman and Bradburn 1982). Sudman and Bradburn (1982) have also suggested that questionnaires between 12 - 16 pages can still obtain reasonable response rates if the population is well-educated and the topic is highly salient. Consequently, despite design improvements, and the advantage of a well-educated population (Postcensal sample members must have at least a bachelors degree), and a questionnaire of moderate salience, the impact of increased questionnaire length and burden on response rates was a concern.

Research has shown that incentives can be effective for increasing response rates (Ferber and Sudman 1974, Armstrong 1975, Kanuk and Berensen 1975, Linsky 1975, Heberlein and Baumgartner 1978, Hansen 1980, Yu and Cooper 1983, Fox, Crask, and Kim 1988, James and Bolstein 1990, James and Bolstein 1992). Few government-sponsored studies, however, have used incentives. As stated by Groves (1989) and others, this may stem from a belief that responding to government surveys is part of one's "civic duty." Thus, some members of the design team expressed concern that this perception of "civic duty," especially in a study conducted by the Bureau of the Census, would cause some Postcensal sample members to become disgruntled over the use of federal dollars for incentives, thus depressing rather than enhancing response rates. After weighing the realities of the increased questionnaire length with the potential advantages and disadvantages of incentives, the decision was made to test the impact of questionnaire length and use of incentives in the 1992 Postcensal pretest.

Methodology

The 1992 Postcensal sample consisted of four treatment groups:

- **Long Form/No Incentive** - A 20-page booklet (12 pages of questions)
- **Mid-Length Form/No Incentive** - a 16-page booklet (7 pages of questions)
- **Long Form/Incentive** - Same 20-page booklet mailed with a $5 check
- **Screener/Follow-up (No Incentive)** - A 6-page screener (3-pages of questions) with a 16-page follow-up questionnaire for those sample members identified as S&E. The follow-up response is not addressed in this paper.

Features of the Pretest design included:

- A 3-year old retired Current Population Survey sample of 3,200 persons. All sample members had a bachelor's degree or higher, and approximately half had been identified as having worked in an S&E occupation.
- With the sample stratified by S&E status, the 3200 sample members were randomly assigned to one of the four treatment groups.
- The same initial data collection procedures were used on all four groups--a pre-notification letter, a first mailing, a reminder letter, and a second mailing.
- The $5 incentive check was included in only the first mailing to the Long Form with Incentive group. Sample members were not reminded of the incentive, nor was the incentive offered during the second mailing, or during the telephone or personal visit follow-ups.
- The three most promising treatment group approaches were selected after the mail phase for telephone and personal visits follow-ups.
- The follow-up sample in the selected groups consisted of a random sample of two-thirds of the mail nonrespondents.

After the first mailing, the 16.5 percent of the questionnaires returned as "undeliverable" were removed from the experiment. The undeliverable questionnaires were, instead, assigned to a separate personal follow-up subsample where only the long form was administered. As expected, given random assignment, the percent of undeliverable questionnaires varied little among the four treatment groups.

Because the "undeliverable" questionnaires were effectively removed from the experiment, we deleted them from our calculations. Thus, for the purposes of this paper, we will report a "cooperation rate" rather than a response rate. We are defining the cooperation rate as:

- Numerator = completed questionnaires from both eligible and ineligible sample members, and "incompletes" (questionnaires missing one or more of the four "critical items")

- Denominator = the original sample minus the undeliverables (all questionnaires presumed to have been received at the first mailing)

We included the ineligibles because the majority came from respondents who had completed questionnaires, but did not have bachelor degrees.

Cooperation Rates

First Mailing. As shown in Table 1, the overall first mailing cooperation rate was 62.2 percent. The Screener and Long Form with Incentive groups performed best, with virtually identical cooperation rates at 67.7 and 67.8 percent, respectively. The Long Form without Incentive group was low at 53.9 percent, 13 percentage points behind the two leaders. The mid-length group had a cooperation rate midway between the two extremes at 59.3 percent.

Second Mailing. The first mailing was followed by a remainder/thank you letter and later, a second questionnaire mailing. The reminder letter and second mailing increased the overall cooperation rate by nearly 9 percentage points. In the three no incentive groups, cooperation rate gains were inversely related to questionnaire length. The Screener group had the largest gain (11.7 percentage points), with slightly more modest gains in the Mid-length and Long Form groups, 9.3 and 8.7 percentage points, respectively. Of the four groups, the Long Form with incentive had the smallest gain--5.5 percentage points.

Cumulative Mail. At the conclusion of the mail phase cooperation rates ranged from a high of 79.4 percent (Screener) to a low of 62.7 percent (Long Form-No Incentive), a difference of nearly 17 percentage points. Again, in the three no incentive groups, cooperation rates were inversely related to questionnaire length. Comparing the two long form groups, using the incentive increased the overall mail response by nearly 11 percentage points.

Telephone. For subsequent follow-up by telephone and personal visits, two-thirds of the mail nonrespondents were sampled in each of the groups, with the exception of the Mid-Length group. The Mid-Length group was dropped because its cooperation rate did not compensate for the analytic limitations imposed by the loss of five pages of survey questions. Although telephone follow-up was initiated in the Screener/Follow-up group, we do not currently have accurate estimates of the cooperation rate for this group.

TABLE 1

MAIL COOPERATION RATES

		First Mailing	Second Mailing	Total Mail
Group	N	Percent	Percent	Percent
Long Form No Incentive	675	53.9	8.7	62.7
Mid-Length	681	59.3	9.3	68.6
Long Form Incentive	674	67.8	5.5	73.3
Screener	674	67.7	11.7	79.4
TOTAL	2704	62.2	8.8	71.0

NOTE: First Mailing $X^2 = 39.57$ $p < .001$

Second Mailing $X^2 = 16.54$ $p < .001$

Total Mail $X^2 = 49.37$ $p < .001$

Telephone follow-up began approximately 8 weeks after the first questionnaire mailing. Because time restrictions precluded any locating beyond using directory assistance, interviews were only completed with those mail nonrespondents who could be easily located. As shown in Table 2, interviews were only completed with 29.9 percent of the telephone sample--31.3 percent of the no incentive and 28.0 percent of the incentive group, a difference of 3.3 percentage points. Although the difference is small, and not statistically significant, it is going in the "wrong" direction. Having received the incentive in the first mailing, we had hypothesized that the sense of social obligation that should have been evoked by norm of reciprocity or social exchange theory (Dillman 1978) would have caused the incentive group to outperform the no incentive group.

Why was the incentive not more effective? Assuming the incentive had been received, it should have maximized completed interviews among contacted sample members. If, however, locating problems predominated, these overall gains would be minimized. This is exactly what we found. Of the 166 no incentive group nonrespondents sampled follow-up, telephone contact was made with about half (49 percent) and interviews were completed with half of those contacted. Among the 118 incentive group nonrespondents sampled, telephone contact was made with only 37 percent, but interviews were completed with nearly two-thirds of those contacted. As shown in Table 2, the incentive group also had a refusal rate that was less than half that of the no incentive group. Thus, the incentive did increase cooperation, but its overall effectiveness was hindered by the greater prevalence of locating problems in the incentive group. This points to the importance of good locating procedures. The 11 percentage point cooperation rate

advantage in the incentive group at the end of the mail phase would suggest that residual nonrespondents would present greater locating problems when an incentive is used.

TABLE 2

TELEPHONE COOPERATION RATES

		Completed Interviews	Refusals
Group	N	Percent	Percent
Long Form No Incentive	116	31.3	16.3
Long Form Incentive	118	28.0	7.6
TOTAL	284	29.9	12.7

NOTE: Completed interviews: $t = -0.602$

Refusals: $t = -2.311$ $p < .05$

Personal Visits. Members of the follow-up sample who could not be contacted by telephone were sent to the field for personal visit follow-ups. Personal visits began in late September, approximately three months after the second questionnaire mailing, the last potential point of contact with these sample members. As shown in Table 3, interviews were completed with about half of these nonrespondents (55.8 percent)--63.2 percent of the no incentive and 47.4 percent of the incentive group, a difference statistically significant at $p < .05$. While we had expected that cooperation rate differences between the two groups would continue to diminish as the number of follow-up increased, the magnitude of the difference during the personal visit phase was surprising. Although it had been four months since the first mailing, we had still expected the principles of the social exchange theory to be in operation.

We looked again at the cooperation rate among those contacted and the refusal rate. As in the telephone follow-up, the contact rate favored the no incentive group (68 percent versus 59 percent), but the cooperation rate was reversed, higher in the no incentive group by almost 12 percentage points. Although not statistically significant, refusals were also slightly lower in the no incentive group. The incentive no longer seemed effective at increasing cooperation.

We looked at one final factor. To be potentially effective, an incentive must be received. As noted earlier, the incentive was not mentioned or offered during telephone or personal visit follow-ups. Given that nonrespondents contacted during personal visits also represented a residual of locating problems, we questioned the extent to which the incentive had been received. We can provide a rough estimate of this number. Sample members interviewed during the personal visit phase were asked what they had done with the last NSCG questionnaire they received. Altogether, 40 percent of the incentive group respondents

believe they had never received the questionnaire, and another 12 percent could not remember or did not know if they had received the questionnaire. Thus, it would appear that no more than half even remembered receiving a questionnaire.

TABLE 3

PERSONAL INTERVIEW COOPERATION RATES

		Completed Interviews	Refusals
Group	N	Percent	Percent
Long Form No Incentive	87	63.2	4.6
Long Form Incentive	76	47.4	11.8
TOTAL	163	55.8	8.0

NOTE: Completed interviews: $t = -2.048$ $p < .05$

Refusals: $t = -1.66$ NS

Total Cooperation Rate. By adjusting the number of completed telephone and personal interviews by a factor of 1.5 (the inverse of the probability of selection into the telephone sample), an estimated combined mail/telephone/personal visit cooperation rate is shown in Table 4 for the incentive and no incentive groups. After two mailings, a reminder postcard, and telephone and personal visit follow-up, only 2 percentage points separate the incentive and no incentive group.

S&E Bias

NSF estimated that a 15-20 percentage point response rate differential existed in the 1982 Postcensal between those trained or employed in S&E fields and those not trained or employed in S&E fields. Since controlling S&E bias is an important NSF goal, we looked at the cooperation rates for the mail returns between CPS sample members who had been identified a priori as having or not having been employed in an S&E field.

With respect to potential S&E bias, all four groups received an advance letter on NSF letterhead. Although the same core questions were used on all four groups, the long form had additional items that focused primarily on previous work experiences. This was a potential source of bias since respondents were required to self-code their occupation using a job list that had many more detailed codes for S&E occupations. The screener form had an additional advantage over the mid-length and long form questionnaires. Rather than self-coding occupation or degree field using lists with more detailed codes for

S&E occupations or degree fields, the screener asked for occupation and degree field using an open-ended format. Based on these differences between the four groups, we expected:

- Questionnaire length to be positively associated with S&E bias where an incentive was not used.

- S&E bias on the screener to be even smaller than would be expected due to questionnaire length alone since respondents were not being asked to use the detailed job and education lists

- The incentive would mitigate the impact of length such that the Long Form with Incentive would have a lower S&E bias than either the Long Form without Incentive or the Mid-length questionnaire.

We found support for all three hypotheses. As shown in Table 5, the S&E and non-S&E cooperation rate differences for the three no incentive groups ranged from 9.8 percentage points in the Long Form-No Incentive ($p < .01$) to 4.0 percentage points for the Screener group (p = NS), with the Mid-length group falling between the two at 8.6 percentage points ($p < .05$). Not only was the difference lowest in the screener group, it was the only group where the difference was not statistically significant.

TABLE 4

COMBINED MAIL/TELEPHONE/PERSONAL VISIT COOPERATION RATES

	Mail	Telephone	Personal Visit	Total[1]
Group	Percent	Percent[a]	Percent[a]	Percent[b]
Long Form No Incentive	62.7	11.6	12.2	86.4
Long Form Incentive	73.3	7.3	8.0	88.4
TOTAL	68.0	9.4	10.1	87.6

[1] Mail: $t = 4.099$ $p < .001$
Telephone: $t = 2.166$ $p < .05$
Personal Visit: $t = -2.096$ NS
Total: $t = .9084$ NS

[a] A 2/3 subsampling rate was used in the follow-up phases. In order to approximate significance tests for returns through the telephone and personal-visit phases, the initial base of attempts was reduced by one-third.

[b] The reported number of completed interviews through all phases was calculated by eliminating 1/3 of the mail completes to reflect the subsampling utilized for the follow-up phases.

TABLE 5

DIFFERENCES IN S&E AND NON-S&E MAIL RESPONSE RATES

	S & E	Non - S & E	Difference
Group	Percent	Percent	Percent
Long Form No Incentive	67.6	57.8	9.8
Mid-Length	72.9	64.3	8.6
Long Form Incentive	76.9	69.8	7.1
Screener	81.4	77.4	4.0
TOTAL	74.7	67.3	7.4

Note: Long Form No Incentive $t = -2.646$ $p < .01$
Mid-Length $t = -2.429$ $p < .05$
Long Form Incentive $t = -2.092$ $p < .05$
Screener $t = -1.286$ NS

Data Quality

To assess data quality, we looked at two factors, the number of "incompletes" (completed questionnaires missing one of the four questions designated as "critical") divided by the total number of complete and incomplete questionnaires and the mean item nonresponse rate. The results are shown in Table 6.

Differences between the four groups were not statistically significant for either factor. Only 3 percentage points separated the four groups with respect to the number of "incompletes" received and less than one percentage point separated the groups on mean item nonresponse. Thus, it appears data quality was not affected by either questionnaire length or the use of an incentive.

TABLE 6

DATA QUALITY

	Quality Measure	
	Incomplete	Mean Item Nonresponse
Group	Percent	Percent
Long Form No Incentive	8.0	2.4
Mid-Length	11.2	2.5
Long Form Incentive	10.8	2.1
Screener	10.7	2.8
TOTAL	10.3	2.5

Incompletes: $X^2 = 2.936$ NS at $p < .05$

Commitment

At the end of each questionnaire, respondents were asked to provide an evening and daytime telephone number for future contact if we needed to "clarify some of the information" provided. The next, and last, item in the questionnaire told respondents that they might be resurveyed again in 1995. They were asked to provide the names, addresses and telephone numbers of two "contact" persons for future locating purposes.

Based on the assumption that offering an incentive increases the importance of a survey in the eyes of the respondent (Goetz et al, 1984, Heberlein and Baumgartner, 1978), we hypothesized that sample members who received the incentive would attach more importance to the study and, consequently, would be more willing to provide future locating information. Table 7 presents preliminary data on the percentage of mail respondents who failed to provide either an evening or daytime telephone number (column 1), no first contact person information (column 2) or no second contact person information (column 3).

Although the item nonresponse for telephone numbers, as shown in column 1, did not vary significantly among the four treatment groups, a statistically significant difference (p. < .05) existed between the long form groups with and without an incentive. Whereas 9 percent of the long form-no incentive respondents failed to provide a telephone number, this dropped to 5.3 percent in the long form incentive group.

TABLE 7

PROVIDING FOLLOW-UP CONTACT INFORMATION[1]

		No Telephone #	No 1st Contact	No 2nd Contact
Group	N	Percent	Percent	Percent
Long Form No Incentive	410	9.0	22.7	37.1
Mid-Length	417	6.2	23.7	37.6
Long Form Incentive	540	5.3	14.6	30.6
Screener	484	6.2	27.4	44.4
TOTAL	1851	6.5	21.8	37.2

Telephone Number: $X^2 = 5.542$ NS

No First Contact: $X^2 = 26.377$ $p < .001$

No Second Contact: $X^2 = 21.040$ $p < .001$

While most respondents were also willing to provide the names of contact persons, they were far less willing to provide this information as compared to providing a current telephone number. As questionnaire length increased, the willingness to provide the name of a contact person also increased, with respondents in the incentive group being the most willing to provide future contact information ($p < .001$). This suggests that questionnaire length influenced the importance that respondents attributed to a study and also supports Goetz's suggestion that providing an incentive also increases the perceived importance of a study.

Discussion

With respect to the cooperation rate, our findings suggest:

- Without an incentive, questionnaire length is inversely related to the cooperation rate

- For a mail survey with one mailing, using an incentive appropriate to the task can significantly increase the cooperation rate. In this instance the cooperation rate for a 20-page questionnaire equalled that of a 6-page questionnaire, and was 13 percentage points higher than the same questionnaire without an incentive.

- With additional mail follow-ups, some of the incentive's power is lost. Here, where additional mail follow-ups included a reminder letter and a second mailing, the Long Form with Incentive lost ground to the brevity of the 6-page questionnaire, but the Long Form with

Incentive still achieved a cooperation rate that was significantly higher than the Long Form with No Incentive (11 percentage points).

- When comparing two questionnaires that both can be considered "long" (the Mid-Length and the Long Form), offering the incentive caused the longer questionnaire to outperform the relatively shorter instrument, even after the mail follow-ups.

With respect to the other factors where we compared the impact of questionnaire length and the use of incentives we found:

- Data quality was not affected by either questionnaire length or use of an incentive

- S&E bias was only marginally reduced by using incentives. While questionnaire length appeared to be positively associated with S&E bias, this finding is confounded by the fact that increased length added questions more likely to generate S&E bias

- Commitment to being contacted for future follow-ups increased with increased questionnaire length and by using incentives, with incentives having the largest impact.

With respect to concerns that using incentives might negatively impact cooperation rates because sample members would be offended about being paid for a task perceived as "civic duty," no apparent offense was noted. None of the sample members who called the number provided on the questionnaire cover letters or reminder letter complained about the incentive; none of the questionnaires were returned with angry notes; and none of the sample members contacted by telephone or personal visits raised this issue. As noted earlier, however, the incentive was only included in the first mailing, so this limited usage might have affected this outcome.

Summary

Questionnaire length is still an important issue for mail surveys, although this pretest has demonstrated what Sudman and Bradburn (1982) and others have suggested; a well-designed mail survey with a burdensome questionnaire can achieve reasonable cooperation rates in a highly educated population--without an incentive. Very brief questionnaires, however, do appear most effective at achieving significantly higher cooperation rates.

Few social policy studies, however, can meet their data needs with a four or six-page questionnaire. Using an incentive can significantly increase respondent cooperation in mail surveys with a fairly burdensome questionnaire and limited follow-ups. The Postcensal pretest also indicated what previous research by James and Bolstein (1990), Nederhof (1983), Heberlein and Baumgartner (1978) and others has shown, that an incentive's power to motivate participation diminishes with successive follow-ups.

Theoretically, however, it seems the motivation that the incentive initially provided should not be so easily lost. How it was applied may be the cause. In this experiment, as well as in the research reported above, the incentive was only used in the first mailing. Because few mail surveys have mailing addresses which are 100 percent current, this may be severely limiting the potential usefulness of the incentive. Experiments are needed where incentives are re-offered during the telephone or personal visit follow-ups of mail surveys to sample members who say they never received the first mailing. Even if the mailing

addresses are current, we need to examine the impact of reminding respondents about the incentive they received. Both of these actions should potentially maximize the incentive's ability to evoke the principals of the norm of reciprocity or the social exchange theory.